

Title (en)

EFFICIENT DATA PROCESSING TO IDENTIFY INFORMATION AND REFORMAT DATA FILES, AND APPLICATIONS THEREOF

Title (de)

EFFIZIENTE DATENVERARBEITUNG ZUR IDENTIFIZIERUNG VON INFORMATIONEN UND NEUFORMATIERUNG VON DATEIEN SOWIE ANWENDUNGEN DAFÜR

Title (fr)

TRAITEMENT EFFICACE DE DONNÉES SERVANT À IDENTIFIER DES INFORMATIONS ET À REFORMATER DES FICHIERS DE DONNÉES, ET APPLICATIONS ASSOCIÉES

Publication

**EP 4052119 A4 20231018 (EN)**

Application

**EP 20882233 A 20201030**

Priority

- US 201916668565 A 20191030
- US 2020058200 W 20201030

Abstract (en)

[origin: US2021133769A1] The present disclosure is directed to systems and methods for identifying demographic information in a data file. The method may include: receiving the data file containing a plurality of fields of demographic information from a third-party, the data file having inconsistent or mislabeled nomenclatures for one or more fields of the plurality of fields or spurious demographic information; analyzing the data file using a machine learning model trained according to other data files to distinguish between each of the plurality of fields of demographic information, the machine learning model being based on a plurality of machine learning algorithms to identify different types demographic information; generating a score indicating a probability that each of the plurality of fields of demographic information was identified correctly; and generating a revised data file labeling each of the plurality of fields of demographic information based on the identified type.

IPC 8 full level

**G06F 7/00** (2006.01); **G06Q 10/10** (2023.01); **G06Q 30/02** (2023.01); **H04N 21/24** (2011.01)

CPC (source: EP US)

**G06F 16/951** (2018.12 - US); **G06F 40/279** (2020.01 - EP); **G06F 40/295** (2020.01 - US); **G06N 20/00** (2018.12 - EP US); **G06Q 10/10** (2013.01 - EP); **G06Q 30/0201** (2013.01 - EP US)

Citation (search report)

- [I] US 2019113905 A1 20190418 - CORR MICHAEL G [US], et al
- [I] WO 2007149216 A2 20071227 - INFORMATION EXTRACTION SYSTEMS [US], et al
- See references of WO 2021087254A1

Designated contracting state (EPC)

AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

DOCDB simple family (publication)

**US 2021133769 A1 20210506**; CN 114830079 A 20220729; EP 4052119 A1 20220907; EP 4052119 A4 20231018; US 2021174380 A1 20210610; WO 2021087254 A1 20210506

DOCDB simple family (application)

**US 201916668565 A 20191030**; CN 202080076168 A 20201030; EP 20882233 A 20201030; US 2020058200 W 20201030; US 202117181519 A 20210222